R4

06/01

PCT/SE 99/01024

S

REC'D 03 AUG 1999
WIPO PCT

PRV

PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen

Intyg
Certificate

Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.

(71) Sökande / Applicant (s) — *Telefonaktiebolaget L M Ericsson, Stockholm SE*

(21) Patentansökningsnummer / Patent application number — *9802193-4*

09/719709

(86) Ingivningsdatum / Date of filing — *1998-06-18*

Stockholm, 1999-07-20

För Patent- och registreringsverket
For the Patent- and Registration Office

Evy Morin

Evy Morin

Avgift
Fee

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

PATENT- OCH REGISTRERINGSVERKET
SWEDEN

Postadress/Adress: Box 5055, S-102 42 STOCKHOLM
Telefon/Phone: +46 8 782 25 00, Vx 08-782 25 00
Telex: 17978, PATOREG S
Telefax: +46 8 666 02 86, 08-666 02 86

METHOD AND APPARATUS IN TRANSMISSION OF IMAGES

1. PATENT PROPOSAL ON INDEPENDENT CODING OF REGIONS OF INTEREST.

2. TECHNICAL FIELD

The present invention relates to a method and a device for coding and extraction of regions of interest (ROI) in transmission of still images and video. The method and the device are particularly well suited for transform based coders like Wavelets and DCT.

3. BACKGROUND OF THE INVENTION AND PRIOR ART

In transmission of digitized still images from a transmitter to a receiver, the image is usually coded in order to reduce the amount of bits required for transmitting the image.

The reason for reducing the amount of bits is usually that the capacity of the channel used is limited. A digitized image, however, consists of a very large number of bits. When transmitting such an image consisting of a very large number of bits over a channel which has a limited bandwidth transmission times for most applications become unacceptably long, if every bit of the image has to be transmitted.

Therefore, much research efforts in recent years have concerned coding methods and techniques for digitized images, aiming at reducing the number of bits necessary to transmit.

These methods can be divided into two groups:

Lossless methods, i.e. methods exploiting the redundancy in the image in such a manner that the image can be reconstructed by the receiver without any loss of information.

Lossy methods, i.e. methods exploiting the fact that all bits are not equally important to the receiver, hence the received image is not identical to the original, but looks, e.g. for the human eye, sufficiently alike the original image.

4. PROBLEM AREA

In some applications parts of a transmitted image may be more interesting than the rest of the image and a better visual quality of these parts of the image is therefore desired. Such a part is usually termed region of interest (ROI). An application in which this can be useful is for example medical data bases or transferring of satelite images. In some cases it is also desired or required that the region of interest is transmitted lossless, while the quality of the rest of the image is of less importance. There is also cases where it is required that the regions of interest is extracted and from the bit stream and decoded without having to decode the whole image.

5. BEST MODE OF THE INVENTION

Below the method for wavelet based coders is described. However, the method is equally applied to DCT- based scehemes.

5.1 Introduction

The basic idea is, when the transformation of the image is done, to use a mask in the transform domain that describes what coefficients in the transform domain are needed for reconsrction of regions of interest and the background in order to *classify* the transform coefficients into segments. The mask can be created by for example use the scheme presented in PO8512 Swedish patent application no. 9703690-9 or P08983SE, Swedish patent application no. 9800088-8. The numbers PO... refers to intenal references at Telefonaktiebolaget L M

A *segment* is here defined as all coefficients in the transform domain that belong to a certain object or the background. The segment can then be divided further into subsets.
A *subset* is here defined as a number of coefficients in a part of the transform domain (e.g. a subband in the wavlet transform case) which is needed for reconstruction and belonging to a segment in the digitized image, see Figure 3.

When this classification is made the segments are coded independently to different levels of accuracy which will yield a bit stream for every segment. These bit streams are then mixed together in som

5.2 Description of operation (Details)

The proposed method works in the *encoder* as follows on a digitized image, see Figure 1:

1. Perform a **forward transformation** on the image to be transmitted.
2. When the information about how to divide the digitized image into objects and background, **a mask is created**, by using for example the technique described in PO8512 and PO8983SE, in the transform domain, describing which coefficients are needed to reconstruct the different objects or the background.
3. Use the mask, **classify** the transform coefficients **into segments.**
4. **Code the segments independently**. This will give the number of bits required for each subset.
5. **Concatenate the bit streams together with stream and header information needed.** This require a bit stream description which can be found below.
6. **Send** the concatenated bit stream.

The method makes it possible for the receiver to have random access to those parts in the image that is needed, see Figure 2, since the information on where in the stream to find the different parts is known.

At the decoder one way of working could be, see Figure 2:

1. **Receive the bit stream** and **decode the header information** needed.
2. **Find** and **decode the segment information** needed.
3. **Create a mask**, by using for example the technique described in PO8512 and PO8983SE, in the transform domain, describing which coefficients are needed to reconstruct the wanted objects or the background.
4. **Decode** the needed **segment data** from the bit stream.
5. **Reconstruct** the needed **segments.**
6. Show image.

Bit Stream Description

In this section the components of a bit stream that could be needed for the technique is presented.

Data Structures and Pointers

- Pointer
  A *pointer* is a set of symbols that defines the position of a bit or byte in a bit stream or a file. Many ways of describing pointers have been defined in computer science. Any suitable such method can be used here. A pointer can be implicitly defined by a specific bit stream composition rule. A pointer can be defined relative an explicitly or implicitly defined

position. A simple way of defining a pointer is to define the number of bits between the required position and a known reference point such as e.g. the first bit in the bit stream.

- Topology description
  The *topology descriptor*, TOP, is a set of symbols that defines the topological relation between the objects and the enumeration of the objects and shapes. This is illustrated in Figure MJ1 where four objects O1, O2, O3, O4 and four shapes S1, S2, S3 and S4 are shown. The topology of the image can e.g. be represented as a tree graph as shown in Fig. MJ2. The nodes and edges of the tree graph can be coded in a data structure using well known methods. p_TOP is a pointer to a topology descriptor

- Shape description
  A *shape descriptor*, S_i, defines the shape of a closed boundary of an object. The shape number, i, is given by a topology descriptor. Many different shape coding techniques can be used. Examples of such methods are chain coding [REF] and the shape coding method in MPEG-4 [REF]. Shape descriptors can be decoded independently once their position in the bit stream is known. p_S_i is a pointer to a shape descriptor.

- Segment description
  An *segment descriptor*, T_i, is a compressed set of symbols that encodes an segment as described above. The segment contains an ordered set of subsets. The object number, i, is given by a topology descriptor. p_T_i is a pointer to a segment descriptor.

- Subset description
  A *subset descriptor*, B_{ij}, is an independently decodable subset, j, of a segment descriptor, T_i, that describes e.g. the coefficients that belongs to a given subband, j, as described above. p_B_{ij} is a pointer to a subset descriptor.

- Multiplexed segment description
  Several *segment descriptions*, $\{T_i, T_j, T_k \ldots\}$, can be *multiplexed* into a joint data structure MT(i, j, k). This is typically done for the purpose of simultaneous progressive transmission of a set of objects. The data structure, MT, is called a multiplexed segment descriptor. Several multiplexing methods can be used. p_ MT is a pointer to a multiplexed segment descriptor.

  - <u>Segment multiplexing methods</u>
    Examples of multiplexing methods are shown in Fig. 4. A simple method is to interleave the subsets belonging to the component segments so that;

    $$MT(i, j, k) = \{ B_{i0}, B_{j0}, B_{k0}, B_{i1}, B_{j1}, B_{k1}, B_{i2}, B_{j2}, B_{k2}, \ldots \},$$

    Where the order of the symbols corresponds to the order in the bit stream with symbols to the left being sent first. Subsets in a multiplexed stream may be excluded if they are known by the decoder.

Bit stream storage format

In order to achieve random access of any image object the stored bit stream or file structure should include at least the following components:

In the image header if needed:

Topology descriptor: TOP

Pointers to shape descriptors: $\{ p\_S_1,\ p\_S_2,\ \ldots\ p\_S_N \}$

Pointers to segment descriptors: $\{ p\_T_0,\ p\_T_1,\ \ldots\ p\_T_N \}$

Optionally pointers to subset descriptors: for each k = [0,N]

$\{ p\_B_{k0},\ p\_B_{k1},\ \ldots\ p\_B_{kN} \}$

In the body of the stored bit stream if needed:

Shape descriptors: { S_1, S_2, ... S_N }

Segment descriptors: { T_0, T_1, ... T_N }

A group of segment descriptors with index {k, l , m ...} can optionally be replaced with a multiplexed segment descriptor MT(k, l, m...)

N is the number of stored objects. The background is the object with index 0.

Progressive transmission of objects with random access

A server is receiving a request for sending image data to a client. The image is stored at the server in the format described in section. Some of the stored data structures (topological information, shapes, segments and subsets) may already have been sent to the receiving terminal. This section describes a procedure for composing a bit stream at the server that is serving the request.

Examples:

Client request

A primitive request contains the following information:

Send objects with numbers k, l, m ... to accuracy n_k, n_l, n_m respectively, where the accuracy is the index of the highest subset that will be sent for each index.

Several primitive requests can be sent. They will be served in the order that they are received or in an order that is otherwise specified.

Procedure for serving a request (details)

Send topological information if needed. TOP is sent in response to the first request of information about an image.

Send all shapes that are necessary for describing the boundary of the requested objects. Shapes that already are known to the decoder need not to be sent. Using the topological tree in figure MJ2 we find that all shapes on the same branch as the object at the same or a lower hierarchical level need to be sent. The server knows the state of the decoder and will only send shapes that not are known at the decoder.

Send (multiplexed) subset descriptors that describes the requested objects to the defined accuracy. Subset descriptors that already are known to the decoder need not to be sent. The client knows e.g. the subsets { B_{k0}, B_{k1}, B_{k2}, B_{k3}} of segment k. The subset descriptors

{ B_{k5}, B_{k6}, B_{k7}} need to be sent if object k is requested to accuracy 7.

5.3 Examples

In this section some examples of situations where the proposed method can be used is explained.

Assume that there is a region in the middle of the image that has the shape of a circle that has to have a better quality than the region outside the circle, henceforth called the background. Both the background and the region should however be transmitted simultaneously. The following then takes place:

- ✓ The original image is transformed with a wavlet transform.
- ✓ A mask in the transform domain is then created. This mask describes which coefficients is needed in the transform domain in order to reconstruct the region and the background. The created mask is then used to classify the coefficients in the transform domain into two segments. One for the region and one for the background.
  The two segments is built up by a number of subsets. The number of subsets are, in this

example, the same as the number of subbands in the transform domain. Thus present situation is then:

- For the region segment: $\{\{r_{0,1}, r_{0,2}, \ldots, r_{0,i}\}, \ldots, \{r_{no\_subbands,1}, r_{no\_subbands,2}, \ldots, r_{no\_subbands,j}\}\}$ where i,j are the number of coeffcients in the different subsets.

- For the background segment: $\{\{b_{0,1}, b_{0,2}, \ldots, b_{0,p}\}, \ldots, \{b_{no\_subbands,1}, b_{no\_subbands,2}, \ldots, b_{no\_subbands,q}\}\}$ where p,q are the number of coeffcients in the different subsets.

✓ The two segments are then coded yielding the following:

- For the region segment:
  A shape descriptor S_r and a segment descriptor $T_r=\{B_{r,0}, B_{r,1}, \ldots, B_{r,no\_subbands}\}$ and a set of subset pointers $\{p\_B_{r,0}, p\_B_{r,1}, \ldots, p\_B_{r,no\_subbands}\}$.

- For the background segment:
  A segment descriptor $T_b=\{B_{b,0}, B_{b,1}, \ldots, B_{b,no\_subbands}\}$ and a set of subset pointers $\{p\_B_{b,0}, p\_B_{b,1}, \ldots, p\_B_{b,no\_subbands}\}$.

✓ The two segments are then mixed together into a single bitstream as follows:
<image header><TOP>< S_r><{$p\_B_{b,0}$, $p\_B_{r,0}$, $p\_B_{b,1}$, $p\_B_{r,1}$, …, $p\_B_{b,no\_subbands}$, $p\_B_{r,no\_subbands}$}><MT(b,r)= {$B_{b,0}$, $B_{r,0}$, $B_{b,1}$, $B_{r,1}$, …, $B_{b,no\_subbands}$, $B_{r,no\_subbands}$}>
In this case the subsets is mixed as shown in the top part of Figure 4.
Note that in the case where the receiver knows the order of how the different parts of the image is sent, the TOP field is not needed.
The first part of the array, from <image header> to ...p_B...}> is in other words a definition of where the different image regions are placed in the rest of the compressed bit stream <MT(b,r)={...B...}>.

✓ The mixed bit stream is then sent to the receiver.

At the decoder side the following will happen:

✓ The image header togther with the topology, shape information and the pointers will be read.

✓ The decoder can now create the same mask as above.

✓ The decoder creates the segments with the underlying subsets.

✓ The decoder starts decoding the mixed bit stream and fills in the transmitted transform coefficients in the corresponding subsets.

✓ A inverse transform is applied.

✓ The image is sent and reconstructed.

This is one way of using the proposed method. Other ways can be to mix the bit streams in a different way. For example, the region can be transmitted first followed by the background. Another example can be that that there are more than one region where they are mixed in a number of ways.

5.4 Advantages

The proposed method has the following advantages:

- taking care of multiple regions of interest.
- classifying the coefficients into segments that is coded independently.
- possible to only send the shape information when needed.
- the ability to have random access in the bit stream to the parts of the image which is in some way vital to the user without having to decode the whole image.

CLAIMS

1. A method in transmission of an image between a transmitter and a receiver, the method comprising the steps of:

- partitioning of the image into at least two image regions; and
- coding the image regions into a coded symbol stream, the coding using a symbolic representation and having predetermined levels of accuracy in the image regions; and
- compressing the coded symbol stream into a compressed bit stream,

characterized in that the method comprises the steps of:

- generating a definition of the different image regions in the compressed bit sream;
- transmitting said definition to the receiver;
- transmitting the compressed bit stream to the receiver; and
- decoding in the receiver predetermined parts of the compressed bit stream with the aid of said definition.

2. An apparatus for transmission of an image comprising:

- a transmitter and a receiver;
- means for patitioning of the image into at least two image regions;

- a coding device for coding the image regions into a coded symbol stream, the coding device using a symbolic representation and having predetermined levels of accuracy in the regions;
- a compressing device for compressing the coded symbol stream into a compressed bit stream; and
- means in the transmitter for transmitting said compressed bit stream to the receiver,

characterized in that the apparatus also comprises:

- means for generating a definition of the different image regions in the compressed bit stream;
- means in the transmitter for transmitting said definition to the receiver;
- decoder in the receiver for decoding predetermined parts of the compressed bit stream with the aid of said definition.

9802193-4

SAMMANDRAG

En bild (3), som föreligger i digitaliserad form, skall överföras på en kanal mellan en sändare och en mottagare. Kanalen har begränsad bandbredd och bilden har dels en mindre viktig bakgrund (R1), dels områden av särskild vikt, intresseregioner (R2, Rn). Bilden transformeras till transformkoefficienter och komprimeras (21) och en mask, svarande mot regionerna (R1, R2, Rn), definieras i transformdomänen (22). Transformkoefficienterna klassificeras (23) och hänföres enligt maskdefinitionen till olika segment (SG1,SG2,SGn). Dessa kodas (24) oberoende av varandra till olika grad av exakthet beroende på hur viktig motsvarande region (R1, R2, Rn) i bilden (3) är. Kodningen ger delbitströmmar (25) vilka länkas samman (26) med bildhuvudinformation (271,272) till en bitström (27) som sändes till mottagaren. Denne avkodar bildhuvudet och segmentinformationen samt återskapar masken i transformdomänen, innefattande form och lägen på regionerna (R1, R2, Rn). Bilden återskapas sedan med hjälp därav till önskad noggrannhet i respsektive region. Flera regioner (R2,Rn) med olika grader av bildkvalité kan definieras och endast intressanta delar av bilden behöver avkodas.

Publiceringsfigur: Figur 2

- the receiever does not even have to receive the whole bit stream belonging to a certain region or the background if a progressive coding method is used.

6. FIGURES

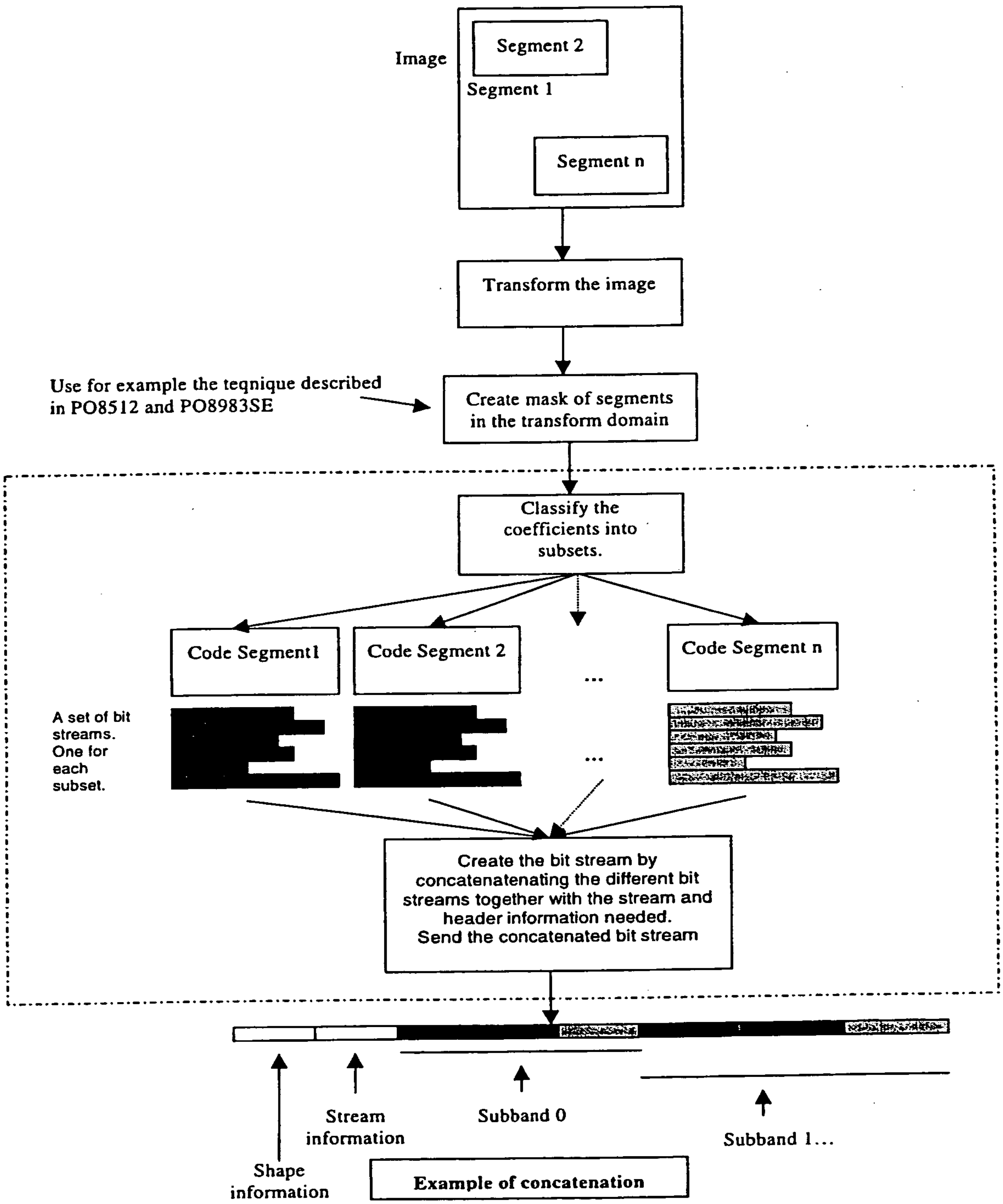


Figure 1: The method of how to code different regions in the encoder.

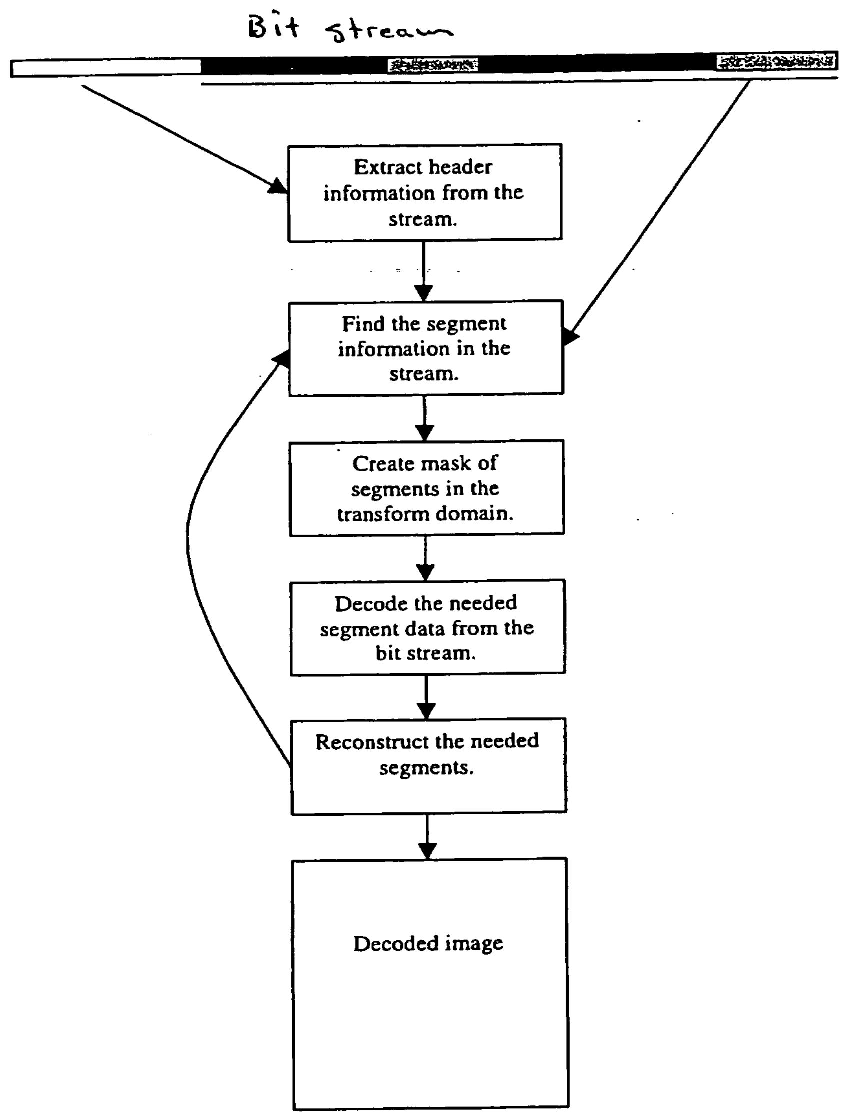
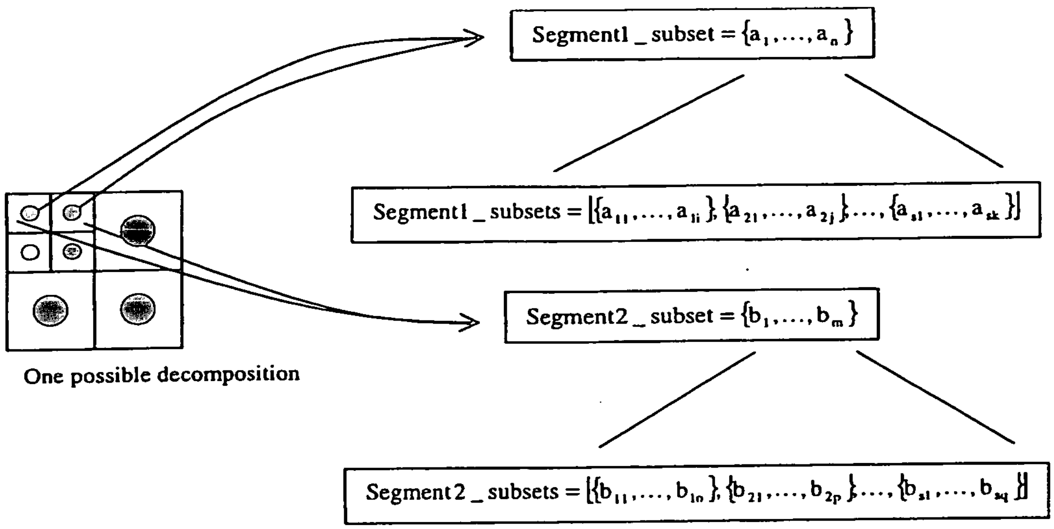


Figure 2: A method for extracting the needed segments from the bit stream.



where n= the number of transform coefficients belonging to segment 1, $n=i+j+k$, m= the number of transform coefficients belonging to segment 2 and $m=o+p+q$.

Figure 3: Classification of transform coefficients into subsets in the case of two segments.

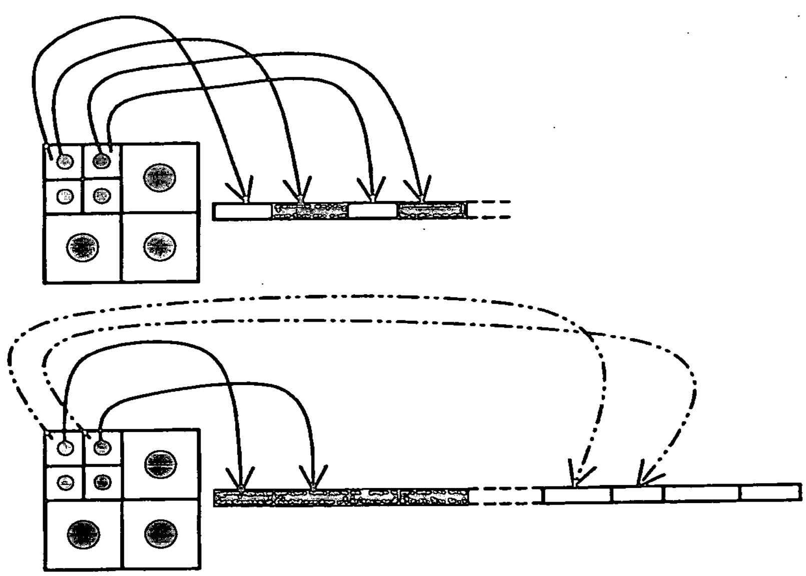


Figure 4: Different ways of concatenating the segments into the bitstream.

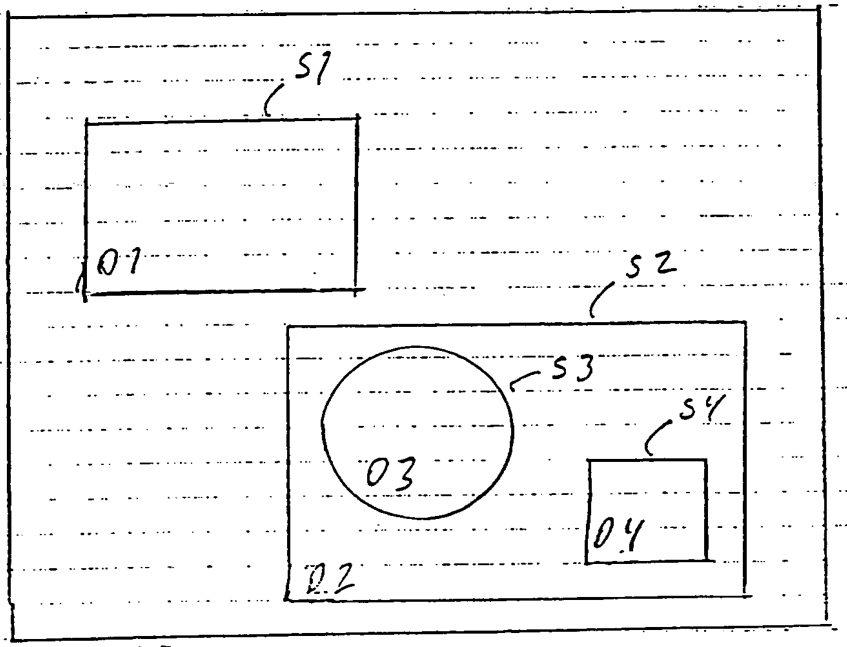


Figure 1: Image with objects

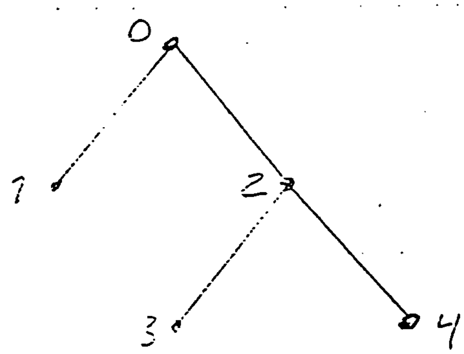


Figure 2: Graphical representation of the topology of the image in fig. 1.